

ABSTRACT

Automatic detection and tracking of multiple individuals includes receiving a frame of video and/or audio content and identifying a candidate area for a new face region in the frame. One or more hierarchical verification levels are used to verify whether a human face is in the candidate area, and an indication made that the candidate area includes a face if the one or more hierarchical verification levels verify that a human face is in the candidate area. A plurality of audio and/or video cues are used to track each verified face in the video content from frame to frame.